Assessing Product Design Using Photos and Real Products

Seong-Eun Moon
Yonsei University
Incheon, Korea
se.moon@yonsei.ac.kr

Jun-Hyuk Kim
Yonsei University
Incheon, Korea
junhyuk.kim@yonsei.ac.kr

Sun-Wook Kim
Hyundai Motor Company R&D Center
Hwaseong, Korea
swkim@hyundai.com

Jong-Seok Lee
Yonsei University
Incheon, Korea
jong-seok.lee@yonsei.ac.kr

CHI'17 Extended Abstracts, May 06–11, 2017, Denver, CO, USA

DOI: http://dx.doi.org/10.1145/3027063.3053336

Abstract
Many studies examining user responses to the product design have been implemented using product photos instead of real products due to practical limitations. In this study, we investigate the validity of such an approach for a particular case with car design evaluation. We compare users' perceptual responses for photos and real cars. In particular, we employ both explicit and implicit response channels, i.e., subjective rating, electroencephalography (EEG), and visual attention. The results show that, although largely similar perceptual responses are obtained in the two cases, significant differences are also observed, which need to be considered in evaluation of product design and user experience.

Author Keywords
product design; user response; questionnaire; electroencephalography (EEG); eye-tracking; visual attention

ACM Classification Keywords
H.1.2 [User/Machine Systems]: Human factors

Introduction
When a product is given to a user, the user perceives it and responds to it implicitly or explicitly, which is strongly connected to decision of purchase. In particular, the design of the product significantly influences this process. A tra-

ditional way to evaluate the design of a product is a questionnaire, which asks users to rate one or multiple aspects of the design. Moreover, biophysiological signals, such as neural activities, gaze patterns, and facial expressions, can be also employed to implicitly monitor users' responses to the product design (emotion, engagement, etc.) [1, 8].

Such evaluation studies are often performed using photos of products instead of real products. In [9], for instance, the relationship between the purchase decision and the aesthetic package design was investigated by using rendered photos of various products. In [10], neural responses to color of cars were studied using rendered images. Using photos for user evaluation is mostly because of practical limitations, i.e., it is cost- and space-consuming to prepare real products or their prototypes and to set an experimental environment with real products. In addition, using real products may complicate the experimental procedure.

However, considering that the ultimate goal of the product design assessment is to understand how users respond to given products, it is important to verify whether there exist any perceptual differences between photos and real products. Unfortunately, this has not been adequately investigated previously.

In this paper, we present our study that investigates perceptual responses to car designs using photos and real products. In particular, we employ three different perceptual aspects, i.e., explicit responses with questionnaire, visual attentional responses based on gaze patterns, and neural responses using electroencephalography (EEG).

Experiments

Procedure

Two car models, named car A and car B from different car brands, were considered in our work. Two experiments were conducted, one with photos, and the other with real cars. The same model of car A was used in the two experiments, while only the color of car B was different due to a practical limitation (white in the first experiment and black in the second experiment).

In the first experiment, four subjects (three males and one female) participated. They were instructed to sit on a chair at ease but move as little as possible to prevent artifacts in the recorded EEG and gaze signals. After the EEG recording device and eye-tracking system were equipped and calibrated, a gray screen was shown for five seconds to record the baseline signals. Then, a photo taken from the front view of each car was shown for 20 seconds one by one. A gray screen was shown for five seconds between the two photos. No particular instruction was given to the participants, so the neural and visual responses were obtained from natural perceptual processes.

Twelve subjects (ten males and two females) participated in the second experiment. The subjects who were involved in the first experiment were excluded in the second experiment in order to avoid perceptual biases induced by the memory recall process. The subjects were asked to stand in front of each car and move as little as possible. Baseline signals were obtained while the subjects closed their eyes for five seconds. Then, they watched each car for 20 seconds.

After each experiment, subjective rating was conducted. The subjects were allowed to watch the photos or real cars as much as they wanted for rating.

Measurement

EEG signals were recorded using Emotiv Epoc+, a wireless EEG device having 14 channels. The positions of the electrodes are shown in Figure 1. The sampling rate was 128 Hz.

The eye-tracking data was measured using Dikablis Eye Tracking Glasses, which is a glass-type eye-tracking device that can record the gaze points at a rate of 50 Hz or 60 Hz.

The questionnaire asked the subjects to rate the car design in the three aspects, namely, preference, luxury, and harmony. Preference indicates overall liking or disliking of the car design, which plays a critical role in the decision making process. Luxurious products satisfy a psychological desire of users, therefore, luxury is frequently considered in the car design. Harmony refers to the consistency among designs of different car parts.

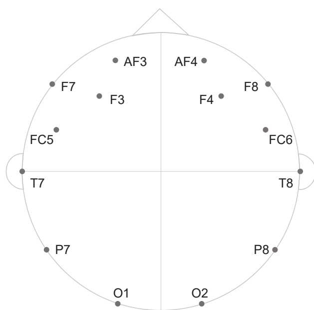


Figure 1: Positions of the EEG electrodes.

The three aspects of the car design were measured with 7-point scoring systems ranging from 1, meaning “strongly dislike” (preference), “very substandard” (luxury), and “very disharmonious” (harmony), to 7, meaning “strongly like” (preference), “very luxurious” (luxury), and “very harmonious” (harmony).

Questionnaire

Results

The results of the subjective rating are summarized in Figure 2 with averages and standard deviations of the scores. In addition, one-tailed t-tests were conducted to examine statistical significance of the differences between the scores obtained using photos and real cars. The difference in the harmony aspect was found to be statistically significant ($p < 0.05$), which is marked in the figure with ‘*’.

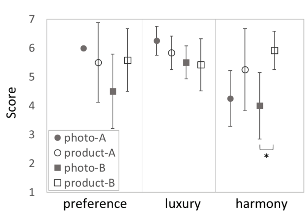


Figure 2: Results of questionnaire. The ‘*’ sign indicates significant difference confirmed by a one-tailed t-test at a significance level of 0.05.

When the two car models are compared, the superiority of one against the other is not consistent in the two experiments. In other words, while car A was always rated higher than car B when their photos were used, car A recorded a higher average score than car B only in terms of luxury when the real cars were used.

The scores obtained with the photos and real cars are also quite different in some cases. In the case of car B, the real car received higher scores than the photo for preference and harmony (with statistical significance for harmony), while they received almost the same average score for luxury. For car A, the photo was rated higher than the real car for preference and luxury, whereas the former received a lower average score than the latter for luxury.

Discussion

The results of the subjective questionnaire indicate that differences of user responses between the photos and real cars are mostly statistically insignificant but still considerable. The largest difference appeared in the case of harmony; for both car models, the photos tend to bias the subjects’ perception in a way that the level of harmony is reduced. This may be due to the fact that the photos weaken depth perception that is important in inducing the sense of three-dimensional harmony. Note that, although the color of car B is different in its photo and real model, the harmony is mostly influenced by the shape rather than the color of the car. The superiority of a car model against the other is also not preserved in the two experiments. These imply that there exist perceptual differences between photo-based assessment and real product-based assessment, which depend on the considered aspect and the product.

EEG

We consider only car A for EEG and gaze pattern analysis, since color difference for car B in the photo and real model may have unwanted influence on neural and attentional responses.

Data Processing
For each electrode, one-second signals were removed both at the beginning and end of the whole signal in order to keep only stabilized responses. To consider temporal changes of perceptual responses, the EEG signals were divided into three-second-long segments without overlapping. The procedure below was conducted for each segment separately.

The segmented EEG signals were bandpass-filtered within 2-64 Hz. The average signal of all electrodes (common average reference) was used for the referencing of the EEG signals. Artifacts due to eye blinking and movement were eliminated by using independent component analysis [6].

The power spectral density (PSD) of the EEG signals, which indicates the level of activation, was employed to compare neural activities. The average PSD across the electrodes was calculated for each of the theta (3-7 Hz), alpha (8-13 Hz), beta (14-29 Hz), and gamma (30-47 Hz) frequency bands, from which the PSD of the corresponding baseline signal was subtracted.

More details of EEG signal processing can be found in [7].

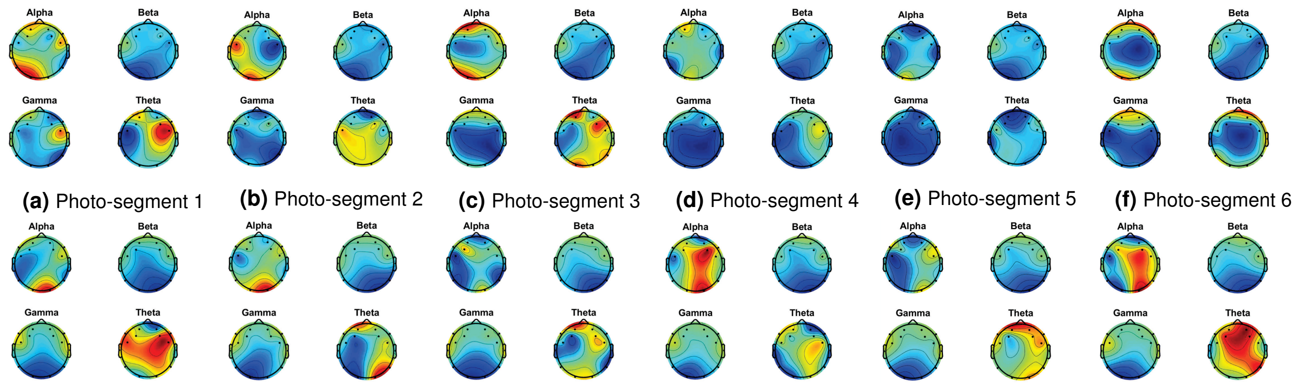
Results
Figure 3 visualizes the level of activation for each frequency band for the two experiments. Segment numbers are mentioned, e.g., segment 1 indicates the first three seconds of the EEG signals, and segment 2 means the next three seconds. The upper row ((a)-(f)) of the figure corresponds to the results of the first experiment, and the lower row ((g)-(l)) is the results of the second experiment. The electrode positions are marked with dots, and the color of the figure indicates the activation level at the corresponding electrode. The area marked by red (blue) color corresponds to an electrode that captured the mostly positively (negatively) activated EEG signal.

	Frequency band				
Segment no.	Alpha	Beta	Gamma	Theta	All
1	0.328	0.642	0.490	0.352	0.562
2	0.146	0.470	0.365	-0.159	0.468
3	-0.020	0.442	0.243	0.440	0.437
4	0.228	0.489	0.399	0.421	0.522
5	0.359	0.536	0.553	0.032	0.503
6	-0.288	0.423	-0.050	-0.096	0.307
All	0.063	0.491	0.297	0.081	

Table 1: Correlation between EEG signal powers of the first and second experiments.

Furthermore, the correlation between the activation patterns obtained from the two experiments was measured in terms of the Pearson linear correlation coefficients between the powers of the two cases (Table 1). The values in the last column of the table indicate the correlation coefficients obtained using all frequency bands, i.e., those between 64-dimensional vectors (14 channels and 4 frequency bands). In addition, the values in the last row were calculated by using 84-dimensional vectors (14 channels and 6 segments)

From Figure 3 and Table 1, the following observations can be made. The beta frequency band shows the highest consistency between the photo-based and real car-based experiments, which is negatively activated in the occipital lobe in both experiments; in Table 1, the correlation coefficients are over 0.4 for all segments, and their average is almost 0.5. The gamma frequency band of the occipital lobe is also negatively activated in the experiment with the real



(a) Photo-segment 1 **(b)** Photo-segment 2 **(c)** Photo-segment 3 **(d)** Photo-segment 4 **(e)** Photo-segment 5 **(f)** Photo-segment 6

(g) Product-segment 1 **(h)** Product-segment 2 **(i)** Product-segment 3 **(j)** Product-segment 4 **(k)** Product-segment 5 **(l)** Product-segment 6

Figure 3: EEG powers in the (a)-(f) first experiment and (g)-(l) second experiment.

car, however, negative activation in the parietal lobe is additionally observed in the experiment with the photo. No significant consistency between the two experiments is observed for the theta and alpha frequency bands, and only partial similarity is observed in an early stage of the experiments. For example, the activation of the alpha and theta frequency bands is found for segments 1 and 3 in the occipital and frontal lobes, respectively.

Table 2 shows the electrodes showing the highest activations. An electrode in the upper row within a cell is the most negatively activated one, and that in the lower row is the most positively activated one. In some cases, no positive activation was observed, for which only the electrodes showing the maximum negative activity were indicated in the table. Electrodes in the columns or rows noted as ‘All’ are the most activated ones among all frequency bands or segments, respectively, where the corresponding frequency bands or segments are indicated in the parentheses.

The most interesting observation in Table 2 is that the EEG activities in the two experiments appear left-right inversely. This is noticeable in the beta frequency band of the occipital lobe, i.e., electrodes O1 and O2 have the maximum negative activity in the two experiments, respectively. In addition, neural activities in the occipital lobe also appear left-right inversely in the alpha frequency band of segments 1, 2, 5, and 6. Similar tendency is shown in a few other cases, e.g., P8 and P7 in the alpha band of segment 1, FC6 and FC5 in the alpha band of segment 2, and F4 and AF3 in the theta band of segment 3.

Discussion
The results show that there exists significant correlation between the EEG signals from the two experiments for the beta and gamma frequency bands. In addition, moderate correlation was also observed for the other frequency bands right after the stimuli were shown. This is probably

	Frequency band									
Segment	**Photo**					**Real product**				
no.	α	β	γ	θ	All	α	β	γ	θ	All
1	P8 O1	O1 -	P8 FC6	FC5 FC6	O1(β) O1(α)	P7 O2	O2 -	O1 T8	AF4 FC6	O2(β) O2(α)
2	FC6 O1	O1 -	P8 T7	AF4 FC5	O1(β) O1(α)	FC5 O2	O2 -	O1 F7	O1 O2	O2(β) O2(α)
3	FC5 O1	O1 -	P8 T8	T7 F4	O1(β) O1(α	FC5 F3	O2 -	O1 T8	P8 AF3	O2(β) T8(γ)
4	P7 AF3	O1 -	FC6 -	T7 FC6	O1(β) AF3(α)	FC5 O2	O2 -	O1 FC5	T7 FC6	O2(β) O2(α)
5	T8 O1	O1 -	P7 -	T7 -	O1(β) O1(α)	FC5 O2	O2 -	O1 T8	F3 AF4	O2(β) AF4(θ)
6	F4 O1	O1 -	P8 AF3	T7 T8	O1(β) O1(α)	O1 O2	O2 -	O1 FC5	T7 AF4	O2(β) O2(α)
All	P7(4) O1(3)	O1(1) -	P8(3) FC6(1)	T7(6) FC6(1)		FC5(2) O2(2)	O2(3) -	O1(6) T8(3)	P8(3) AF4(6)	

Table 2: Most activated EEG electrodes.

because the subjects concentrated on the experiments at the beginning but tended to pay less attention as time went.

Furthermore, it is revealed that the most active electrodes in the two experiments often appear in the opposite sides. In previous studies, it was shown that the left-right EEG asymmetry index is appropriate as a measure of perceptual states such as emotion [3], cognitive processes of verbal and spatial tasks [4], and behavior [2]. Therefore, monitored neural responses for the product design as a perceptual state may become significantly different depending on whether a photo or a real product is used.

Eye-Tracking

Data Processing

First, the effect of head movement was corrected in the recorded gaze signals. Then, the gaze points were classified into fixation and saccade by using the EyeMMV toolbox [5], and only the fixation points were kept.

Four features were obtained from the gaze points: total number of fixations, total duration of fixations, standard deviation of fixation durations, and normalized total length of the scanpath. Here, normalization of the scanpath length was done using the width of the car in the viewpoint of the subjects.

Results

Table 3 compares the four gaze pattern features for the two experiments. Average and standard deviation values across subjects are shown in the table. We performed Wilcoxon two-sample tests in order to examine statistical significance of the differences between the two experiments. As a result, the difference of the total duration of fixations was statistically significant ($p < 0.05$).

Discussion

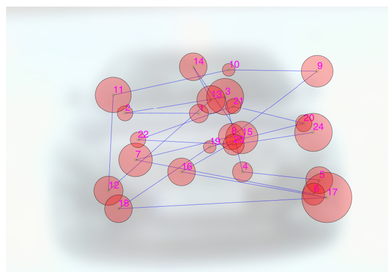
The statistical significance of the difference in the total fixation duration indicates that the subjects attended more on the real car than the photo. Although statistical significance was not found, a smaller number of fixations in the case of the real car also supports this observation.

For qualitative analysis, Figure 4 shows a representative example of the eye-tracking data of the two experiments. The subject concentrated on a smaller number of regions in the case with the real car than in the case with the photo (15 vs. 24) and, therefore, the normalized length of the scanpath became smaller in the former case (3.8 vs. 6.1). In addition, in the case with the real car, the fixation durations differ more significantly depending on the fixation point (0.85 vs. 0.53), which means that the subject tended to spend sufficient time only at selected regions.

These observations imply that the viewing patterns for the real car and the photo are quite different, which would be the reason behind the differences in the subjective ratings

	Features			
Cases	*Total number of fixations*	*Total duration of fixations (s)*	*Std of duration of fixations (s)*	*Length of scanpath*
Photo	19.75 (3.50)	18.04* (0.37)	0.67 (0.12)	6.06 (1.64)
Real Product	18.33 (4.31)	18.79* (0.66)	0.80 (0.26)	5.62 (1.79)

Table 3: Comparison of the eye-tracking features.



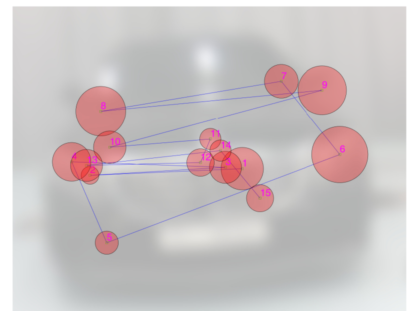


Figure 4: Example in the photo experiment (left) and real product experiment (right). The radius of a circle represents the duration of the corresponding fixation.

and EEG activities. Different levels of sense of reality and immersion in the two cases probably influence on determining such different visual perception.

Conclusion

Using real products in an experiment usually imposes several difficulties. First, it is sometimes hard to secure a space for the experiment. The space needs to be isolated and large enough to implement the experiment smoothly using a number of products. This requirement becomes critical particularly when the product is large, e.g., car as in our case. Second, real products are more expensive in comparison to their photos. Moreover, when the product does not exist, its prototype needs to be made, which is costlier than using existing products. Third, the implicit monitoring of user responses (such as neural and physiological responses) should be implemented more carefully. In our case, we calibrated the eye-tracker and EEG amplifier each time when subjects moved. Forth, it is time-consuming. Movements of products or subjects and frequent calibration require a significant amount of time for the experiment. A lengthy experiment leads to the fatigue of subjects, particularly when they are wearing an eye-tracker or EEG sensors.

Nevertheless, our results demonstrate that using real products is desirable to understand user responses better and more accurately, because the perceptual responses to the photos and real cars were shown to be considerably different. Such differences were observed in all channels of responses, i.e., explicit ratings, neural responses, and viewing patterns. The superiority of one car model against the other in the ratings changed in some cases. The activation levels in the EEG signals were quite similar in both experiments especially at the early stage of the experiments, but the activation asymmetry of the left and right hemispheres was significantly different, indicating that the subjects' perception and internal states may be different. Viewing patterns were also different, showing a higher level of concentration in the case of the real car.

The perceptual differences observed in our study in the two experiments would be mainly due to the limited sense of reality, engagement, and immersion when the photos were used. As an extension of our work, therefore, it would be interesting to examine user responses using more realistic presentation methods without real products, such as 360-degree images, 3D images, and virtual reality.

Acknowledgements
This work was supported by Hyundai Motor Company.

References
[1] Rafael A. Calvo and Sidney D'Mello. 2010. Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing* 1, 1 (2010), 18–37.

[2] James A. Coan and John J. B. Allen. 2002. Frontal EEG asymmetry and the behavioral activation and inhibition systems. *Psychophysioloby* 40, 1 (2002), 106–114.

[3] Richard J. Davidson. 1988. EEG measure of cerebral asymmetry: conceptual and methodological issues. *International Journal of Neuroscience* 39, 1 (1988).

[4] David Galin and Ron R. Ellis. 1975. Asymmetry in evoked potentials as an index of lateralized cognitive processes: relation to EEG alpha asymmetry. *Neuropsychologia* 13, 1 (1975), 45–50.

[5] Vassilios Krassanakis, Vassiliki Filippakopoulou, and Byron Nakos. 2014. EyeMMV toolbox: an eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. *Journal of Eye Movement Research* 7, 1 (2014), 1–10.

[6] Scott Makeig, Anthony J. Bell, Tzyy-Ping Jung, and Terrence J. Sejnowski. 1996. Independent component analysis of electroencephalographic data. In *Proceedings of the Advances in neural information processing systems 8*. 141–151.

[7] Seong-Eun Moon and Jong-Seok Lee. 2015. Perceptual experience analysis for tone-mapped HDR videos based on EEG and peripheral physiological signals. *IEEE Transactions on Autonomous Mental Development* 7, 3 (2015), 236–247.

[8] Ahmed Fazle Rabbi, Abongwa Zony, Pablo de Leon, and Reza Fazel-Rezai. 2012. Mental workload and task engagement evaluation based on changes in electroencephalogram. *Biomedical Engineering Letters* 2, 3 (2012), 139–146.

[9] Martin Reimann, Judith Zaichkowsky, Carolin Neuhaus, Thomas Bender, and Bernd Weber. 2010. Aesthetic package design: a behavioral, neural, and psychological investigation. *Journal of Consumer Psychology* 20, 4 (2010), 431–441.

[10] Hong Wang and Ningning Zhang. 2010. The analysis on vehicle color evoked EEG based on ERP method. In *Proceedings of the International Conference on Bioinformatics and Biomedical Engineering*. 1–3.